UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards (OAQPS)
Research Triangle Park, North Carolina 27711

December 14, 2012

MEMORANDUM

SUBJECT: Re-analysis of the relationship between 24-hour and 4-hour visibility index levels.

FROM: Neil Frank and Mark Schmidt, Air Quality Analysis Group, OAQPS/OAR /s/

TO: PM NAAQS Review Docket (EPA-HQ-OAR-2007-0492)

This memorandum presents an update to the analysis of the relationship between 24-hour and 4-hour estimated light extinction levels presented in Appendix G of the Policy Assessment (PA) for the review of the particulate matter (PM) national ambient air quality standards (NAAQS) (US EPA, 2011a). This update includes (1) a revision to the estimates of 24-hour and 4-hour visibility index levels using a consistent approach to estimate organic mass and (2) a better accounting for the uncertainties in 24-hour and 4-hour estimates of the visibility index values via the use of an alternative statistical approach. The updated analysis reveals that the 24-hour and 4-hour light extinction values are even more closely related than presented previously in the PA or the proposed rule for the PM NAAQS (77 FR 38890, June 29, 2012).

Background

In Appendix G of the PA, EPA explored various approaches for calculating adjusted candidate protection levels (CPLs) for a 24-hour average $PM_{2.5}$ light-extinction indicator that are generally equivalent to CPLs of 20, 25, and 30 deciviews (dv) for a daily maximum daylight 4-hour average $PM_{2.5}$ light extinction indicator, on an aggregate or central tendency basis. In developing these adjusted CPLs, EPA compared, for 15 study areas, (i) values of 24-hour average $PM_{2.5}$ light extinction calculated using the original IMPROVE algorithm combined with site-specific daily data on $PM_{2.5}$ mass and composition and site-specific long-term relative humidity conditions with (ii) values of daily maximum daylight 4-hour average $PM_{2.5}$ light extinction calculated from the hourly results of the Urban-Focused Visibility Assessment (UFVA) (U.S. EPA, 2010b) modeling of hourly $PM_{2.5}$ light extinction. The Appendix G approach for estimating 24-hour light extinction involved estimating organic mass (OM) for each Chemical Speciation Network (CSN) sampling day by subtracting a sampler-dependent estimate of the organic carbon (OC) artifact from the OC measurement to estimate $PM_{2.5}$ OC and multiplying the result by 1.4 (i.e., OM = $PM_{2.5}$ OC *1.4). This was distinctly different than the approach followed in the UFVA, which entailed numerous and complex data processing steps to generate hourly $PM_{2.5}$ composition information from less time resolved data, including application of the SANDWICH approach (Frank, 2006) for estimating OM via material balance.[1]

Some of the approaches described in Appendix G to determine generally equivalent 24-hour adjusted CPLs focused on comparing 24-hour and 4-hour light extinction values in each of the 15 urban areas assessed in the UFVA, whereas other approaches focused on comparisons of aggregated data across the urban areas. Two of these city-specific approaches (regressions of annual 90th percentile light extinction values and regressions of 3-year average light extinction design values) gave nearly identical results and were determined in Appendix G to be most appropriate for identifying generally equivalent 24-hour adjusted CPLs. These approaches identified as A and B (as depicted in Figures G-7 and G-8, Appendix G of the PA) had high R-squared values of the regressions and used data from days with $PM_{2.5}$ light extinction conditions in the range of 20 to 40 dv. In contrast, other approaches were highly influenced by $PM_{2.5}$ light extinction conditions well below this range. Based on these analyses and staff conclusions, including a preference for Approach B, presented in Appendix G of the PA, the EPA identified adjusted 24-hour CPLs of 21, 25, and 28 dv as being generally equivalent to 4-hour CPLs of 20, 25, and 30 dv, as shown below in Table 1.

Other approaches, including Approaches C, D, and E, also shown in Table 1, provided information about how city-specific or pooled 24-hour values compared to 4-hour values. While these approaches had lower R-squared values than Approaches A and B, they illustrated that there is a significant range in values among different cities, suggesting that average or 90th percentile values might not accurately represent all locations. Notably, Approaches C and E generated a range of city-specific estimates of generally equivalent 24-hour levels that encompassed the range of levels considered appropriate for 4-hour CPLs, including the CPL of 30 dv at the upper end of that range.[2]

[1] In Appendix F of the Policy Assessment, the EPA specifically evaluated which multiplier would produce 24-hour results most similar to the SANDWICH approach using 24-hour $PM_{2.5}$ organic carbon derived from the new Chemical Speciation Network (CSN) carbon monitoring protocol established in 2007, and concluded that a multiplier of 1.6 is most appropriate for purposes of comparing the hourly $PM_{2.5}$ light extinction with calculated 24-hour extinction (see Appendix F, section F.6 for a full explanation).

[2] As discussed in more detail in Appendix G of the PA, some days have higher values for 24-hour average light extinction than for daily maximum 4-hour daylight light extinction, and consequently an adjusted "equivalent" 24-hour CPL can be greater than the original 4-hour CPL. This can happen for two reasons. First, the use of monthly average historical RH data will lead to cases in which the f(RH) values used for the calculation of 24-hour average light extinction are higher than all or some of the four hourly values of f(RH) used to determine daily maximum 4-hour daylight light extinction on the same day. Second, $PM_{2.5}$ concentrations may be greater during non-daylight periods than during daylight hours.

Table 1. Relationships between 24-hour and 4-hour levels from Policy Assessment (Corrected Table G-6)[3]

Approach	Description	24-hour level equivalent to 20dv for 4 hour (range among 15 cities)	24-hour level equivalent to 25dv for 4 hour (range among 15 cities)	24-hour level equivalent to 30dv for 4 hour (range among 15 cities)
A	3-year 90_{th} percentile design values regression	22 dv	25 dv	28 dv
B	Annual 90_{th} percentile values regression	21 dv	25 dv	28 dv
C	All-days city-specific regressions, then averaged	19 dv (17 - 21)	23 dv (21 - 25)	27 dv (24 - 30)
D	All days pooled regression	19 dv	23 dv	27 dv
E	Median ratios, then averaged	19 dv (17 - 21)	24 dv (21 - 26)	29 dv (25 - 31)

Estimating 24-hour and 4-hour light extinction levels using a more consistent approach

The first step in this reanalysis of urban speciation data is to consistently estimate OM with a multiplier of 1.6. As noted above, EPA previously derived 4-hour and 24-hour light extinction levels using different approaches. The 4-hour levels were derived with the UFVA approach which entailed numerous and complex data processing steps to generate hourly $PM_{2.5}$ composition information from less time resolved data, including an estimate of OM using the SANDWICH material balance, while the 24-hour levels were based on the original IMPROVE algorithm and using a 1.4 multiplier applied to $PM_{2.5}$ OC. Based on the results of the analysis presented in Appendix F of the PA, a 1.6 multiplier was found to provide a suitable estimate of OM from $PM_{2.5}$ OC measured by the new CSN monitoring protocol. This produced estimates comparable to OM estimated using the SANDWICH approach.

Revising the statistical model

The relationships between 24-hour light extinction levels and 4-hour light extinction levels derived in the PA were all based on a simple linear regression which utilizes an ordinary least squares (OLS) approach. The OLS approach determines a fitted line that makes the sum of squared vertical distances between the points of the data set and the fitted line as small as possible. This results in an equation of the form $Y = b_0 + b_1X$, where Y is the predicted variable (i.e., 24-hour light extinction levels), and X is the explanatory variable (i.e., 4-hour levels). This approach does not account for any error in X, and as a result may overestimate the intercept “a” and underestimate the slope “b” when there are errors associated with X. When the error in X is large, the biases in the estimated parameters of the regression line can also be large.

[3] Note that the city-specific ranges shown in Table G-6, Appendix G of the Policy Assessment are incorrectly stated for Approaches C and E. Drawing from the more detailed and correct results for Approaches C and E presented in Tables G-7 and G-8, respectively, the city-specific ranges in Table G-6 for Approach C should be 17 – 21 dv for the CPL of 20 dv; 21 – 25 dv for the CPL of 25 dv; and 24 – 30 dv for the CPL of 30 dv; the city-specific ranges in Table G-6 for Approach E should be 17 – 21 dv for the CPL of 20 dv; 21 – 26 dv for the CPL of 25 dv; and 25 – 31 dv for the CPL of 30 dv.

There is an alternative "errors-in-variables" regression model designed to determine the line of best fit by accounting for errors in observations for both X and Y. This model is known as a Deming regression. A special case of the Deming model is an orthogonal regression which assumes that the uncertainties in the observations of X and of Y are approximately equal and determines the line of best fit by minimizing the sum of squared perpendicular distances from the data points to the regression line. The orthogonal regression results in an equation $Y = c_0 + c_1 X$. A Deming regression in general and an orthogonal regression in particular will more appropriately result in a lower intercept and a higher slope than those derived using an OLS regression, and accordingly, provide more accurate values of the predicted variable (Y) for each chosen value of the explanatory variable (X). Another property of orthogonal regression is its symmetry. Unlike OLS, where the regression of Y on X does not produce the same equation as the regression of X on Y, that is not true with orthogonal regression. Using the notation above for orthogonal regression, the relationship between X and Y can be derived directly from the relationship of Y on X. Specifically, this orthogonal regression equation will be $X = (Y - c_0) / c_1$.

<u>Estimating the relationship between 24-hour and 4-hour light extinction levels using a multiplier of 1.6 for OM and alternative regression models</u>

Accordingly, in the tables presented below, we have updated the analyses presented in Appendix G of the PA using a multiplier of 1.6 for OM for organic carbon measured with the new CSN monitoring protocol and using two regression models, including the original OLS regression approach and the orthogonal approach described above. The formulas are:

$Y = b_0 + b_1X$, (based on OLS regression); and

$Y = c_0 + c_1X$, (based on orthogonal regression).

For each model, comparisons between 24-hour and 4-hour values were made based on the following approaches,[4] consistent with Appendix G of the PA:

- Approach A: 3-year 90th percentile design values
- Approach B: Annual 90th percentile values
- Approach C: All-days city-specific values
- Approach D: All-days pooled values for all cities

The results of these revised analyses are shown below in Table 2 for the updated OLS regressions and in Table 3 for the orthogonal regressions, which we consider more appropriate for the reasons discussed above. For Approach B (annual 90th percentile values), Table 3 shows that the 24-hour CPLs of 19, 24, and 29 dv are estimated to be generally equivalent to 4-hour CPLs of 20, 25, and 30 dv. This is in contrast to the values for Approach B of 21, 25, 28 dv presented as comparable to these same 4-hour CPLs in the PA. Thus, compared to the findings in the PA (Appendix G, Table G-6) as corrected and presented above as Table 1, the 24-hour and 4-hour light extinction values are more closely related for the highest CPL based on this improved

[4] Approach E presented in Appendix G of the PA is based on the median values per city and is not affected by the regression analyses. Therefore, the results from Approach E do not change with the reanalysis and are not included in Tables 2 and 3.

re-analysis, which incorporated a more consistent approach to estimating OM in the calculations of 4-hour and 24-hour light extinction values. Moreover, the low and high end of the range of city-specific levels from Approach C using the orthogonal regression method bracket the average levels derived from all approaches. In particular, we note that the high end of the range of city-specific values from Approach C is 30 dv with OLS regression but is 36 dv with orthogonal regression. The updated city-specific results are shown in Table 4. The information from Tables 2-4, along with the parameter estimates of the two regression models, is combined into Table 5.

The revised analysis more clearly indicates that city-specific 24-hour light extinction can be both higher (for the highest CPL) and lower (for the lowest CPL) than 4-hour values (see Table 3, Approach C and Table 4); and that the average values are generally more closely related (Table 3, Approaches A – D) than presented previously in Appendix G of the PA.

Table 2. "Equivalent" Levels for Calculated 24-Hour $PM_{2.5}$ Light Extinction Using Four Approaches, based on Ordinary Least Squares Regression (analogous to Table G-6 from the PA, with update to 1.6 multiplier for 24-hour speciation sampling days with new carbon protocol)

Approach	Description	24-hour level equivalent to 20dv for 4 hour (range among 15 cities)	24-hour level equivalent to 25dv for 4 hour (range among 15 cities)	24-hour level equivalent to 30dv for 4 hour (range among 15 cities)
A	3-year 90th percentile design values regression	22 dv	25 dv	28 dv
B	Annual 90th percentile values regression	21 dv	25 dv	28 dv
C	All-days city-specific regressions, then averaged	20 dv (17-21)	23 dv (21-26)	27 dv (24-30)
D	All days pooled regression	19 dv	23 dv	27 dv

Table 3. "Equivalent" Levels for Calculated 24-Hour $PM_{2.5}$ Light Extinction Using Four Approaches, based on Orthogonal Regression (analogous to Table G-6 from the PA, with update to 1.6 multiplier for 24-hour speciation sampling days with new carbon protocol)

Approach	Description	24-hour level equivalent to 20dv for 4 hour (range among 15 cities)	24-hour level equivalent to 25dv for 4 hour (range among 15 cities)	24-hour level equivalent to 30dv for 4 hour (range among 15 cities)
A	3-year 90th percentile design values regression	22 dv	25 dv	28 dv
B	Annual 90th percentile values regression	19 dv	24 dv	29 dv
C	All-days city-specific regressions, then averaged	20 dv (17-22)	24 dv (22-29)	29 dv (25-36)
D	All days pooled regression	19 dv	24 dv	28 dv

Table 4. Comparison of City Specific Results using OLS and Orthogonal Regressions

City	Approach C: City Specific Results 24-hour levels (in dv) equivalent to 4-hour levels of 20, 25 and 30 dv (by regression model)						
	OLS Regression				Orthogonal Regression		
	20 dv	**25 dv**	**30 dv**		**20 dv**	**25 dv**	**30 dv**
Tacoma	21	26	30		22	29	36
Fresno	19	23	28		18	24	29
Los Angeles	18	22	26		18	22	26
Phoenix	20	24	28		21	27	32
Salt Lake City	17	22	27		18	23	28
Dallas	18	21	24		18	22	25
Houston	20	23	25		20	23	27
St. Louis	20	23	26		20	24	28
Birmingham	21	24	28		21	25	29
Atlanta	20	23	27		19	24	28
Detroit	20	24	28		20	24	28
Pittsburgh	21	25	29		21	25	30
Baltimore	21	24	28		21	25	29
Philadelphia	18	22	26		17	22	27
New York City	21	24	27		20	24	28
Average	20	23	27		20	24	29
Minimum	17	21	24		17	22	25
Maxium	21	26	30		22	29	36

Table 5. Revised Relationships between 24-Hour and 4-Hour Levels Using Multiplier of 1.6 for OM and Two Regression Models (OLS and Orthogonal Regressions)

								Equivalent 24hr level derived from 4hr		
Approach	Metric	Estimation Method	Regression Model	area	intercept	slope	r-square	4 hr = 30 dv	4 hr = 25 dv	4 hr = 20 dv
A	3-year 90th percentile design values regression	simple linear regression (OLS)	DV_pct90_24 = b0 + b1*DV_pct90_4x	U.S.	9.75	0.61	0.80	28	25	22
		orthogonal regression	DV_pct90_24x = c0 + c1*DV_pct90_4	U.S.	8.50	0.66	0.80	28	25	22
B	Annual 90th percentile values regression	simple linear regression (OLS)	pct90_24 = b0 + b1*pct90_4x	U.S.	7.07	0.71	0.71	28	25	21
		orthogonal regression	pct90_24x = c0 + c1*DV_pct90_4	U.S.	0.35	0.95	0.71	29	24	19
C	All-days city-specific regressions, then averaged	simple linear regression (OLS)	Daily_24x = b0 + b1*daily_4	Tacoma	1.50	0.96	0.58	30	26	21
				Fresno	-0.65	0.96	0.83	28	23	19
				Los Angeles	3.26	0.75	0.81	26	22	18
				Phoenix	2.79	0.85	0.62	28	24	20
				Salt Lake City	-1.15	0.93	0.80	27	22	17
				Dallas	5.52	0.62	0.64	24	21	18
				Houston	9.82	0.51	0.45	25	23	20
				St. Louis	6.72	0.66	0.61	26	23	20
				Birmingham	7.20	0.68	0.61	28	24	21
				Atlanta	4.84	0.74	0.62	27	23	20
				Detroit	4.38	0.78	0.82	28	24	20
				Pittsburgh	4.18	0.82	0.80	29	25	21
				Baltimore	6.09	0.74	0.70	28	24	21
				Philadelphia	3.07	0.76	0.57	26	22	18
				New York City	8.11	0.63	0.64	27	24	21
				U.S. average >				**27**	**23**	**20**
C	All-days city-specific regressions, then averaged	orthogonal regression	Daily_24x = c0 + c1*daily_4	Tacoma	-4.70	1.35	0.58	36	29	22
				Fresno	-2.91	1.06	0.83	29	24	18
				Los Angeles	1.50	0.81	0.81	26	22	18
				Phoenix	-0.64	1.10	0.62	32	27	21
				Salt Lake City	-3.15	1.04	0.80	28	23	18
				Dallas	3.27	0.73	0.64	25	22	18
				Houston	6.79	0.67	0.45	27	23	20
				St. Louis	3.57	0.80	0.61	28	24	20
				Birmingham	3.77	0.84	0.61	29	25	21
				Atlanta	0.75	0.92	0.62	28	24	19
				Detroit	2.85	0.85	0.82	28	24	20
				Pittsburgh	2.20	0.92	0.80	30	25	21
				Baltimore	3.62	0.86	0.70	29	25	21
				Philadelphia	-3.24	1.01	0.57	27	22	17
				New York City	5.59	0.73	0.64	28	24	20
				U.S. average >				**28**	**24**	**20**
D	All days pooled regression	simple linear regression (OLS)	Daily_24x = b0 + b1*daily_4	All days pooled regression	3.58	0.79	0.71	27	23	19
		orthogonal regression	Daily_24x = c0 + c1*daily_4	All days pooled regression	0.78	0.92	0.71	28	24	19

References

U.S. EPA (2010b). Particulate Matter Urban-Focused Visibility Assessment – Final Report. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA-452/R-10-004. July 2010. Available: http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_2007_risk.html.

U.S. EPA (2011a). Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA 452/R-11-003. April 2011. Available: http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_2007_pa.html.